AIP Conference Proceedings

Improving ensemble decision tree performance using Adaboost and Bagging

Md. Rajib Hasan, Fadzilah Siraj, and Mohd Shamrie Sainin

Citation: AIP Conference Proceedings **1691**, 030008 (2015); doi: 10.1063/1.4937027
View online: http://dx.doi.org/10.1063/1.4937027
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1691?ver=pdfcov
Published by the AIP Publishing

Improving Ensemble Decision Tree Performance Using Adaboost and Bagging

Md. Rajib Hasan [1,a)], Fadzilah Siraj [2,b)] and Mohd Shamrie Sainin [3,c)]

[1,2,3] *University Utara Malaysia*

[a]Corresponding author: rajib@live.com.my
[b]fadzilahsiraj@gmail.com
[c]shamrie@uum.edu.my

Abstract. Ensemble classifier systems are considered as one of the most promising in medical data classification and the performance of deceision tree classifier can be increased by the ensemble method as it is proven to be better than single classifiers. However, in a ensemble settings the performance depends on the selection of suitable base classifier. This research employed two prominent esemble s namely Adaboost and Bagging with base classifiers such as Random Forest, Random Tree, j48, j48grafts and Logistic Model Regression (LMT) that have been selected independently. The empirical study shows that the performance varries when different base classifiers are selected and even some places overfitting issue also been noted. The evidence shows that ensemble decision tree classfiers using Adaboost and Bagging improves the performance of selected medical data sets.

INTRODUCTION

Ensemble classifier systems are considered as one of the most promising in medical data classification [1]. Estimating classifier accuracy in medical domain is important in that it allows to evaluate how accurately a given classifier will correctly label future medical data, i.e., data on which the classifier has not been trained. Medical data classification is acknowledged as an area of increasing importance, yet also poses many difficulties [1] when different base classifiers are involved in the ensemble classifiers [2]. Data from medical studies typically suffer from either multi modal issue or high dimensionality or missing value [3]. Ensemble methods are therefore suitable to be applied to medical datasets [4] as it has favorable properties to handle these issues [5].

Ensemble methods like Bagging and Boosting which combine the decisions of multiple hypotheses are some of the strongest existing machine learning methods in medical settings. The challenging approaches in ensemble known as Bagging and Boosting is to use different base classifiers that use different training sets [6]. Due to different base classifiers in ensemble, there is a variance between ensemble accuracy [7]. Hence, this research employed different decision tree base classifiers in Adaboost and Bagging to identify the performance of ensemble classifiers in the selected medical data set.

LITERATURE REVIEW

Ensemble methods are considered as a more advanced data mining technique where multiple classifiers (in this study Decision Tree Classifiers abbreviate as "DTC") are combined to produce better predictions and more robust methods [8]. An ensemble classifier is a classifier that combines multiple base classifiers for final classification (see Fig.1). In this research the focus remained on decision tree ensemble.

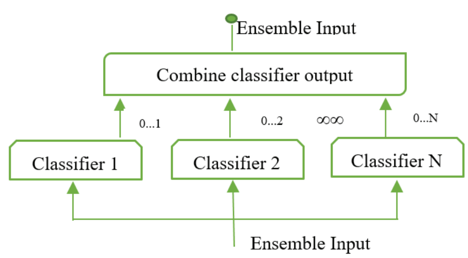

FIGURE 1. Ensemble mechanism

Innovation and Analytics Conference and Exhibition (IACE 2015)
AIP Conf. Proc. 1691, 030008-1–030008-7; doi: 10.1063/1.4937027

Ensemble classifiers refers to the procedures employed to train multiple learning machines and combine their outputs, treating them as a combination of DTC to decision makers [9]. Ensemble Learning like Bagging and Boosting which combine the decisions of multiple hypotheses are some of the strongest existing machine learning methods [10]. An ensemble is itself a supervised learning algorithm, because it can be trained and then used to make predictions [11].

- **Bagging**

Bootstrap aggregating invented by Breiman in 1999 [12]. It is abbreviated as bagging [13] and known as one of the earliest ensemble algorithms [14]. It involves having each method in the ensemble vote with equal weight. In order to promote variance method, bagging trains each method in the ensemble using a randomly drawn subset of the training set [15]. As an example, the random forest algorithm combines random decision trees with bagging to achieve very high classification accuracy [16].

- **Adaboost**

Adaptive Boosting has been introduced by Freund & Schapire in 1996. [17]. It involves incrementally building an ensemble by training each new method instance to emphasize the training instances that previous methods misclassified. In some cases, boosting has been shown to yield better accuracy than bagging, but it also tends to be more likely to over-fit the training data. By far, the most common implementation of Boosting is Adaboost, although some newer algorithms are reported to achieve better results (Darwish, 2013). The boosting algorithm does not create each base classifier independently. Instead, the classifiers are created sequentially where the next base classifier assigns more weights to the mistakes that the previous classifier made and classification is based on weighted base classifiers [15]. Adaptive boosting (Adaboost) is one of the most popular boosting algorithms [18].

METHODOLOGY

In this study, Adaboost and Bagging have been employed with several decision tree base classifiers namely: random forest, random tree, j48, j48 graft and LMT depicted in Fig 2.

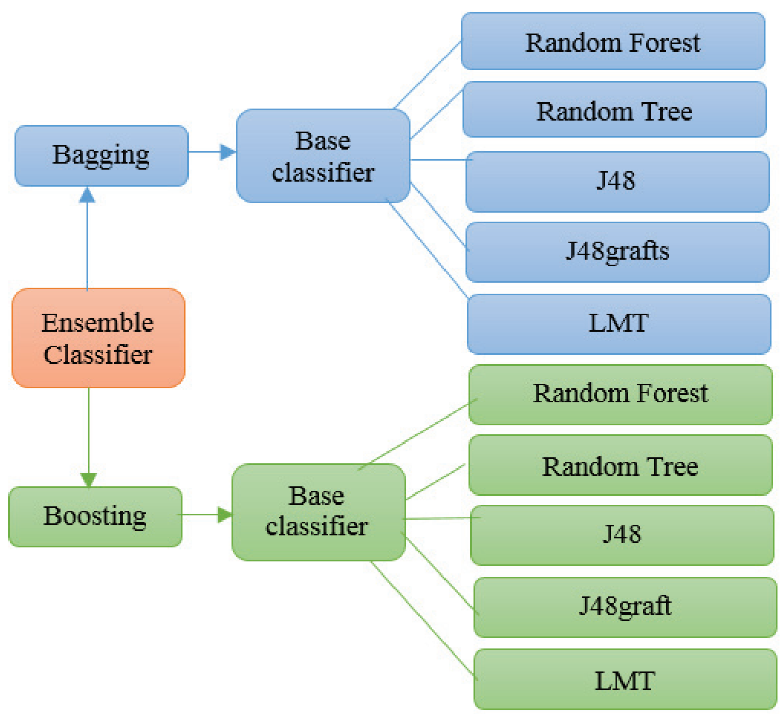

FIGURE 2. Approach

RESULTS

To achieve the objectives, three medical data sets extracted from UCI are employed in the study, namely Hepatitis data, Wisconsin Breast Cancer data and Pima Indian diabetes data. Five suitable classifier namely: random forest, random tree, j48, j48grafts and LMT [2] have been selected independently as the base classifiers in the ensemble methods and the results are presented in the following subsections.

Adaboost

FIVE (5) selected algorithms among the fourteen decision trees are employed in Adaboost and Bagging [2]. These 5 algorithms include random tree, random forest, J48, J48 graft and LMT and the results are shown in Fig. 3. For comparison purposes, each method is compared with a single classifier performance. The results of ADA Boost algorithm for Random Tree are exhibited in Fig 3.

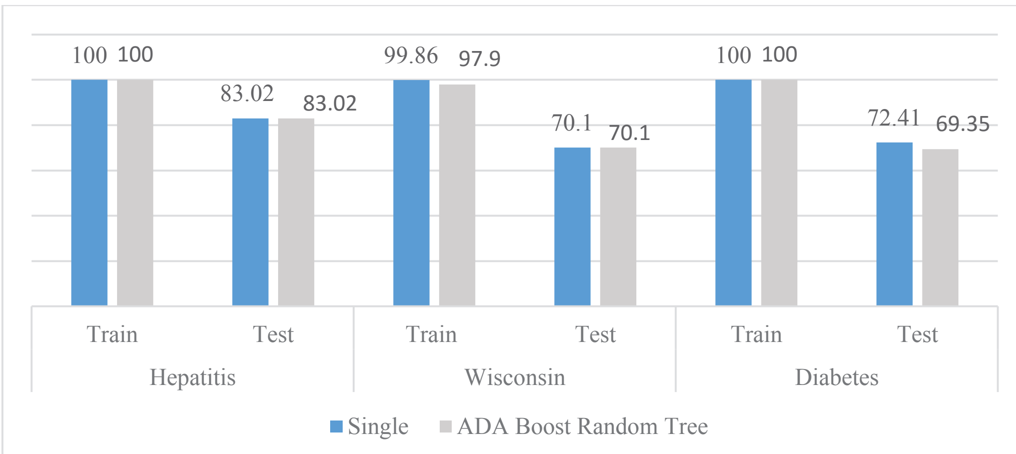

FIGURE 3: Single classifier versus ADA Boost: Random Tree

For hepatitis and Wisconsin datasets, there is no difference between the test results of single classifier and ADA Boost with Random Tree. However, the performance of ADA Boost for diabetes dataset does not increase the test performance. The results indicate that regardless of the data set used in the experiments, Adaboost with Random Tree does not increase the performance of single classifier.

As in the case of ADA Boost with Random Forest for all datasets, the test performance of ADA Boost is lower than single classifier (Fig. 4). Similar observation has been seen when the Random Tree is used in the previous experiment. The largest difference in test performance is the Wisconsin breast cancer dataset with the difference of 6.18% compared with the test performance of single classifier.

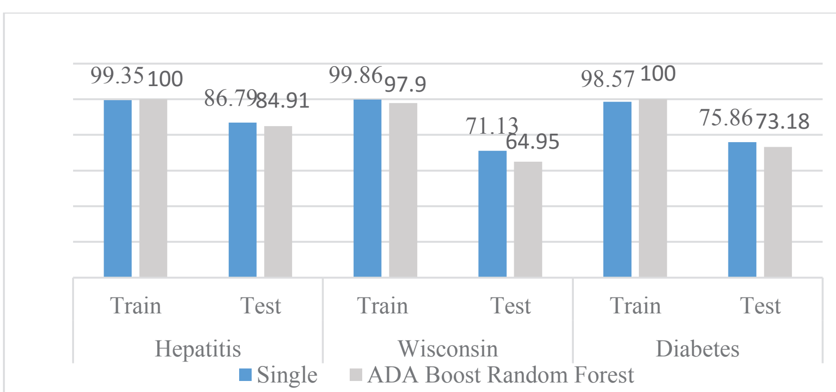

FIGURE 4. Single classifier versus ADA Boost: Random Forest

The overall performance of ADA Boost with J48 does not improve the test performance of single classifier except for Diabetes dataset with an improvement of 1.53% (Fig. 5). Considering the training performance of ADA Boost with J48 versus single classifier, it appears that the training results for ADA Boost for Hepatitis and Diabetes.

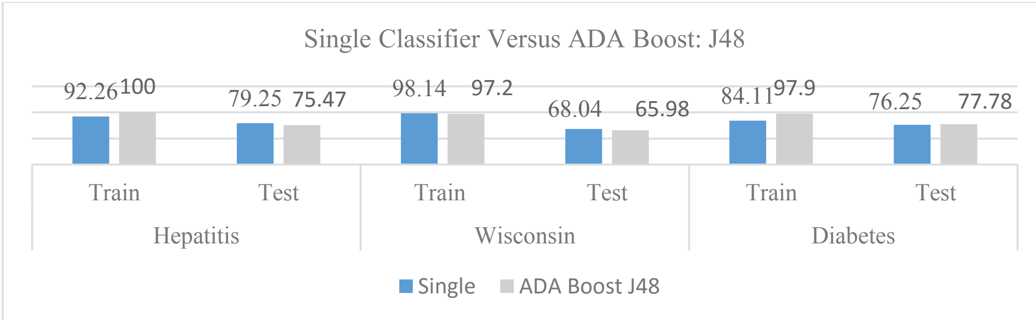

FIGURE 5. Single classifier versus ADA Boost: J48

Similar observation is seen in the training and test performance of ADA Boost with J48 graft. In fact, all test performance of ADA Boost with J48graft are lower than single classifier (see Fig 6).

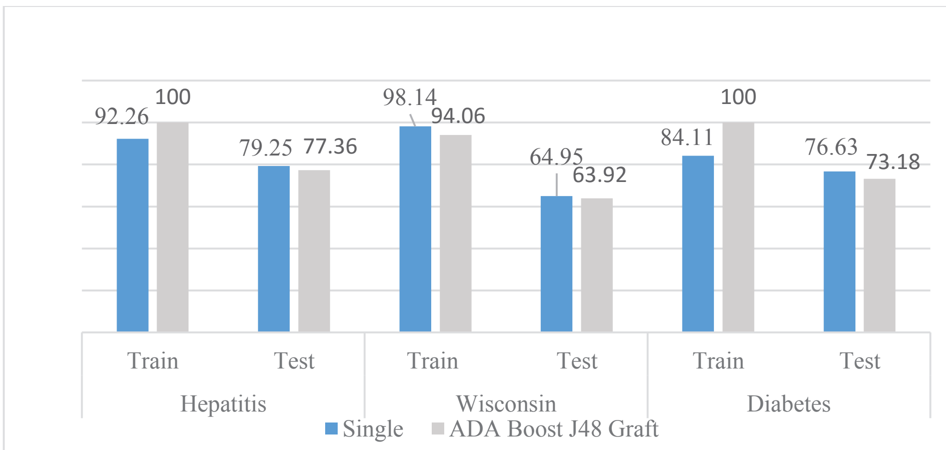

FIGURE 6. Single classifier versus ADA Boost with J48

Unlike other ADA boost ensemble classifiers, ADA Boost with LMT shows the highest (95.79% versus 74.22%) increase in the test performance for Wisconsin breast cancer dataset with an increase of 21.57% (Fig. 7). In addition, all training results are 100% across the tree datasets. However, over-fitting seems to occur for diabetes dataset (100% for training and 71.26% testing).

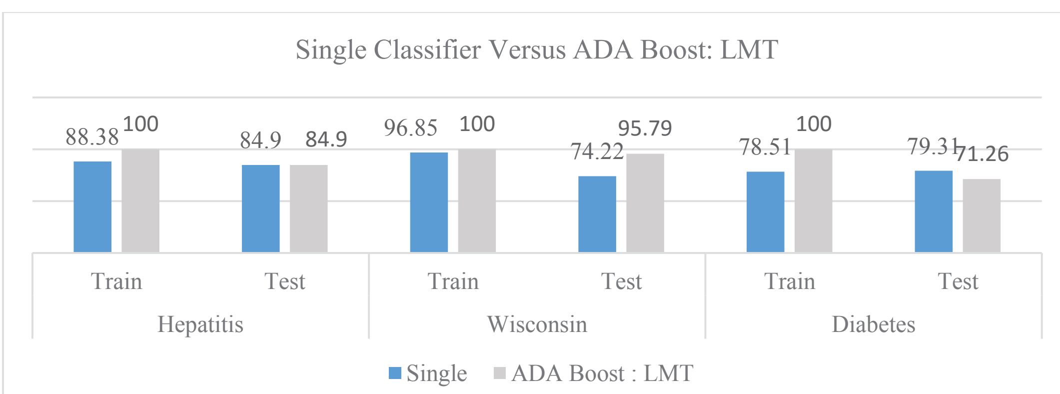

FIGURE 7. Single classifier versus ADA Boost with LMT

Bagging

Similar to ADA Boost, the **FIVE (5)** selected algorithms are also employed in bagging including random tree, random forest, J48, J48 graft and LMT. In contrast with ADA boost results, the test performance of Bagging with Random Tree shows higher performance than single classifier (Fig. 8), in particular hepatitis (86.79%) and diabetes data (77.78%). One interesting point to note that all training performance of Bagging with Random Tree are lower than ADA Boost performance and yet some of the test performance are even higher. Closer inspection of the test performance reveals that the performance for hepatitis data set increases by 3.77%. Furthermore, comparing training and test performance (98.71% versus 86.79%) for this dataset, there is no indication of over-fitting. Hence, in this case, the best test performance obtained is 86.79%.

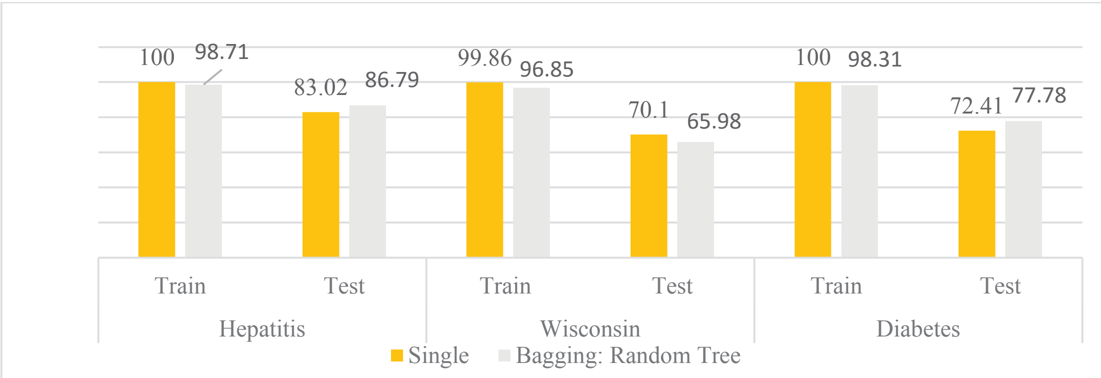

FIGURE 8. Single classifier versus Bagging: Random Tree

The training performance of Bagging with Random Forest shows similar trend (Fig. 9). Although the test result for Diabetes is slightly higher than single classifier, there is an indication of over-fitting in this case. Hence, there is no best test performance for this algorithm.

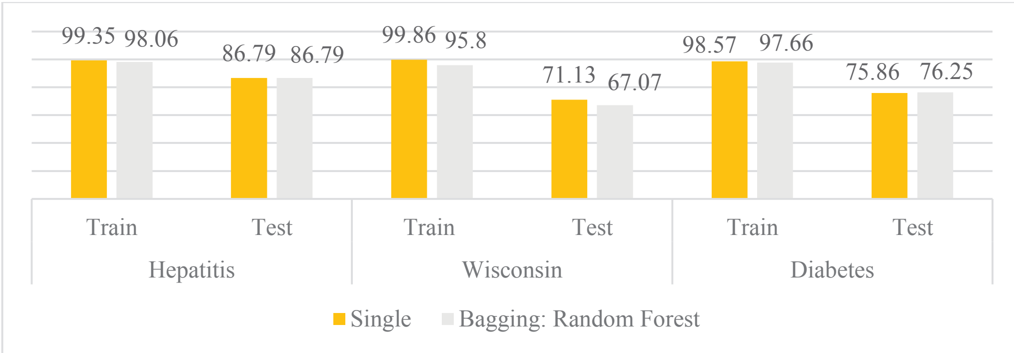

FIGURE 9. Single classifier versus Bagging: Random Forest

The overall results of Bagging with J48 in terms of training and testing do not show a certain trend (Fig. 10). There is no consistency in the performance with the data set except for hepatitis dataset, the test performance obtained is 81.13% with an increase of 1.88%.

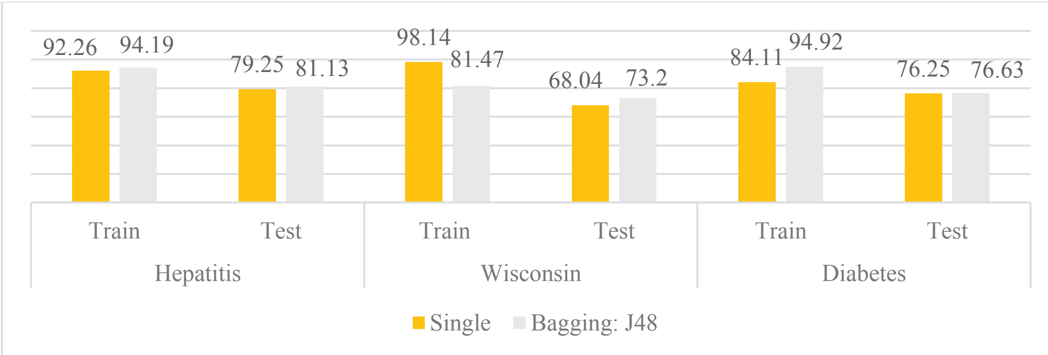

FIGURE 10. Single classifier versus Bagging with J48

Among the THREE (3) data sets for Bagging with J48 Graft, only the test performance using Wisconsin data set shows an increase in the performance (5.15%) as shown in Fig. 11.

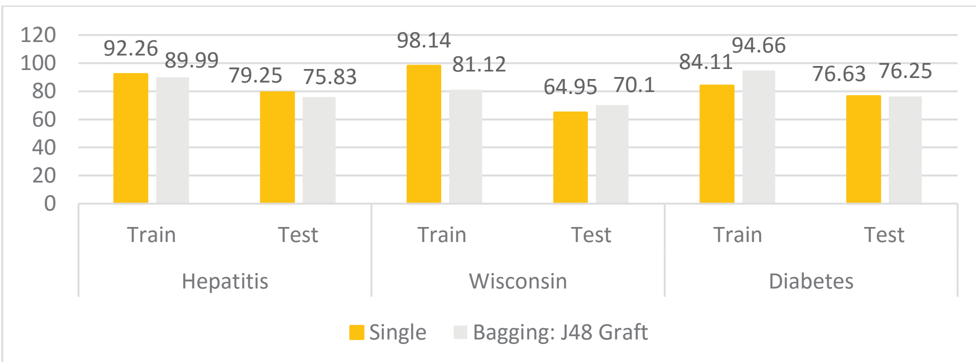

FIGURE 11. Single classifier versus Bagging with J48 Graft

The test performance of Bagging with LMT using TWO (2) data sets, namely the hepatitis and Wisconsin breast cancer show the highest increase so far. For hepatitis and Wisconsin data sets, there is an increase of 5.66% and 21.57% respectively (Fig. 12). So far, the empirical results using LMT have shown the highest test performance as opposed by the single classifier.

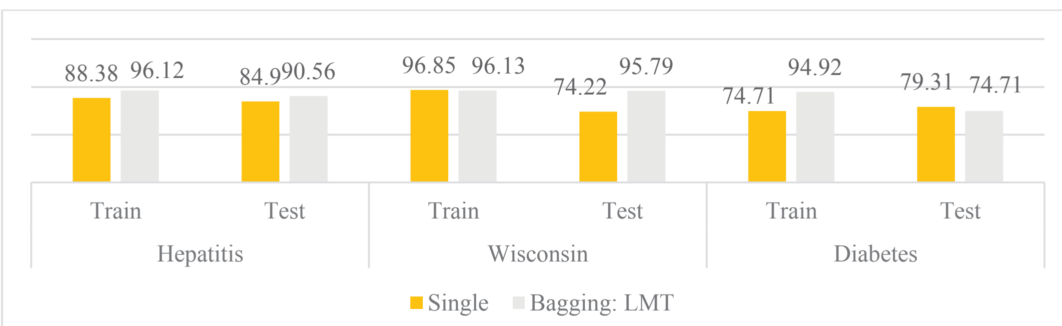

FIGURE 12. Single classifier versus Bagging with LMT

CONCLUSION

Both single and ensemble decision tree classifiers are very popular since they are relatively fast to train [19][20]. Previous empirical results in this study show that some classifiers behave differently with different data sets. Nevertheless, the analysis has shown that ensemble classifiers of decision tree (either boosting or bagging) mostly perform better than single classifier of decision tree. Future

REFERENCES

1. B. Krawczyk and G. Schaefer, "Dealing with the Difficult Learning Situation," *Neural Netw. Appl. Electr. Eng.*, vol. 1, no. 1, pp. 12–15, 2012.
2. M. R. Hasan, N. A. A. Bakar, F. Siraj, M. S. Sainin, and S. Hasan, "Single Decision Tree Classifiers' Accurcay on Medical Data," in *Proceedings of the 5th International Conference on Computing and Informatics, ICOCI, 2015*, 2015, pp. 1–6.
3. L. Nanni, A. Lumini, and S. Brahnam, "A classifier ensemble approach for the missing feature problem.," *Artif. Intell. Med.*, vol. 55, no. 1, pp. 37–50, May 2012.
4. R. Polikar, A. Topalis, D. Parikh, D. Green, J. Frymiare, J. Kounios, and C. M. Clark, "An ensemble based data fusion approach for early diagnosis of Alzheimer's disease," *Inf. Fusion*, vol. 9, no. 1, pp. 83–95, Jan. 2008.
5. H. Moon, H. Ahn, R. L. Kodell, S. Baek, C.-J. Lin, and J. J. Chen, "Ensemble methods for classification of patients for personalized medicine with high-dimensional data.," *Artif. Intell. Med.*, vol. 41, no. 3, pp. 197–207, Nov. 2007.
6. I. Kioutsioukis and S. Galmarini, "*De praeceptis ferendis*: good practice in multi-model ensembles," *Atmos. Chem. Phys. Discuss.*, vol. 14, no. 11, pp. 15803–15865, Jun. 2014.
7. Parvin, M. MirnabiBaboli, and H. Alinejad-Rokny, "Proposing a classifier ensemble framework based on classifier selection and decision tree," *Eng. Appl. Artif. Intell.*, vol. 37, no. 1, pp. 34–42, Jan. 2015.
8. K. Holst and A. Manga, "SAP Predictive Analysis – Real Life Use Case Predicting Who Will Buy Additional Insurance," *SAP COMMUNITY Netw.*, vol. 1, no. 1, pp. 1–104, 2013.
9. G. Brown, "Ensemble Learning," *Encyclopedia of Machine Learning*. Springer Press, pp. 1–24, 2010.
10. P. Melville, "Creating Diverse Ensemble Classifiers," The University of Texas at Austin, 2003.
11. B. A. Fensterstock, J. Salters, and R. Willging, "On the Use of Ensemble Models for Credit Evaluation," *Credit Financ. Manag. Rev.*, vol. 1, no. 1, pp. 1–14, 2013.
12. L. Breiman, "Using adaptive bagging to debias regressions," 1999.
13. M. Kusum and M. Rupali, "A REVIEW ON VARIOUS CLASSIFICATION ALGORITHMS FOR AN INCREMENTAL SPAM FILTER," *Int. J. Appl. or Innov. Eng. Manag.*, vol. 2, no. 11, pp. 325–331, 2013.
14. C. Zhang and Y. Ma, *Ensemble Machine Learning: Methods and Applications*. New York: Springer New York, 2012.
15. D. Che, Q. Liu, K. Rasheed, and X. Tao, "Software Tools and Algorithms for Biological Systems," *Adv. Exp. Med. Biol.*, vol. 696, no. 1, pp. 191–199, 2011.
16. D. Darwish, "Data Mining: Concepts, Models, Methods, and Algorithms," *Int. J. Comput. Sci.*, vol. 10, no. 4, pp. 103–111, 2013.
17. Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," *Mach. Learn. Proc. Thirteen. Int. Conf.*, vol. 1, no. 1, pp. 148–156, 1996.
18. R. E. Schapire, "Explaining AdaBoost," USA, 2013.
19. S. P. Mahila and A. Pradesh, "Ensemble Decision Tree Classifier For Breast Cancer Data," *Int. J. Inf. Technol. Converg. Serv.*, vol. 2, no. 1, pp. 17–24, 2012.
20. B. Chandra, "Heterogeneous Node Split Measure for Decision Tree Construction," *IEEE Int. Conf. Syst. Man, Cybern. (SMC), 2011*, pp. 872–877, 2011.